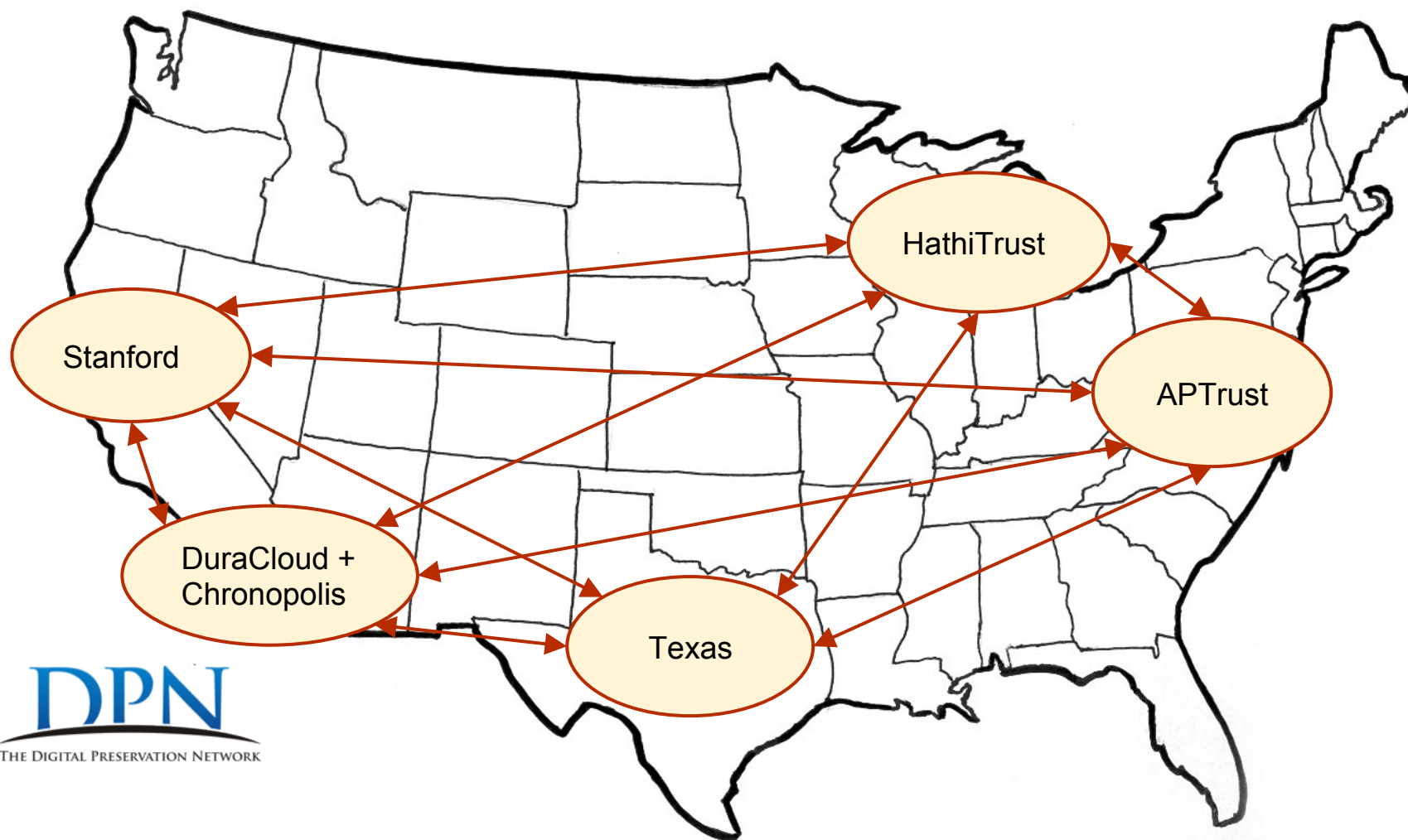
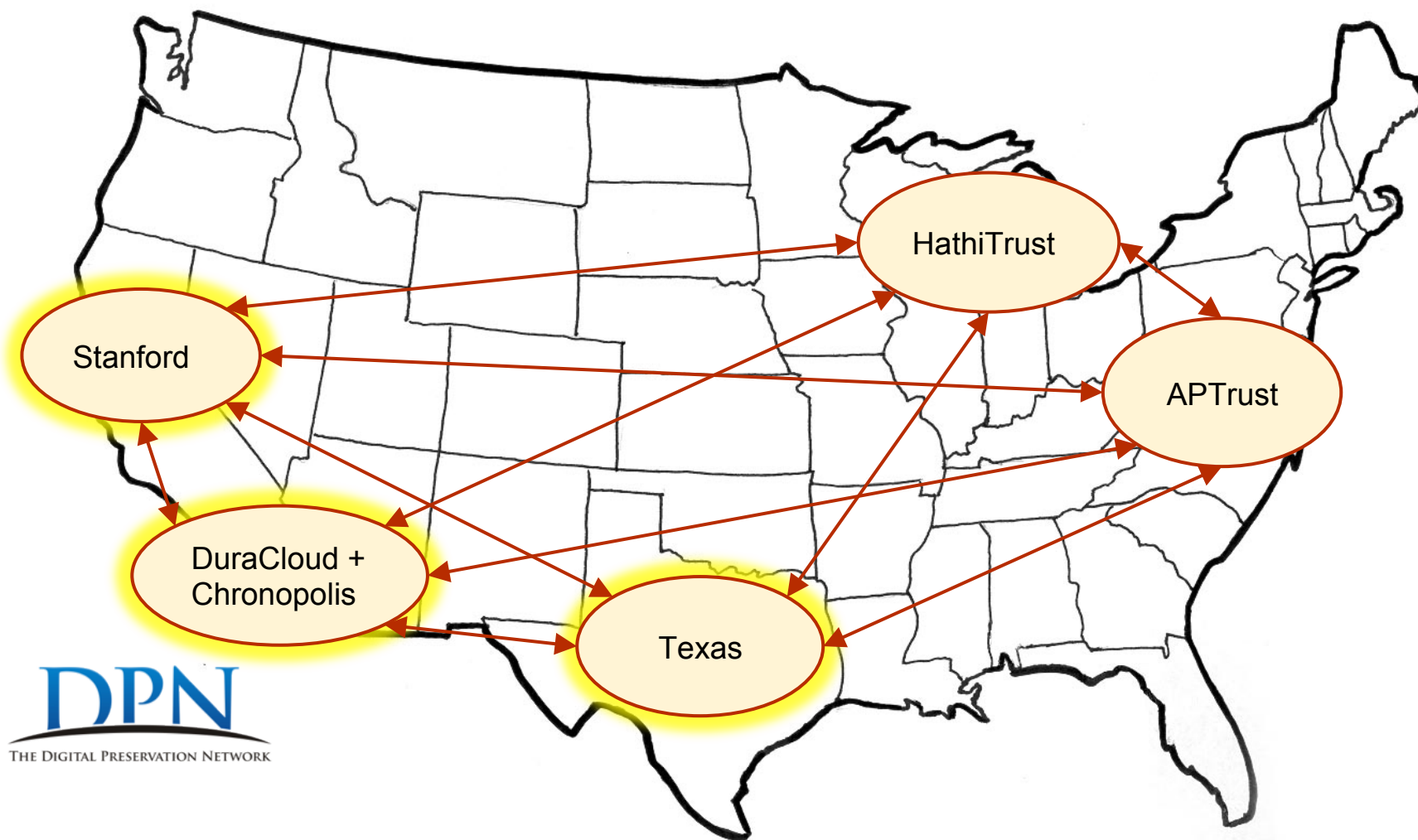




Ingest into the Digital Preservation Network: Standard Pipelines

David Minor - UC San Diego Library
Dan Galewsky - University of Texas Libraries
Tom Cramer - Stanford University Libraries







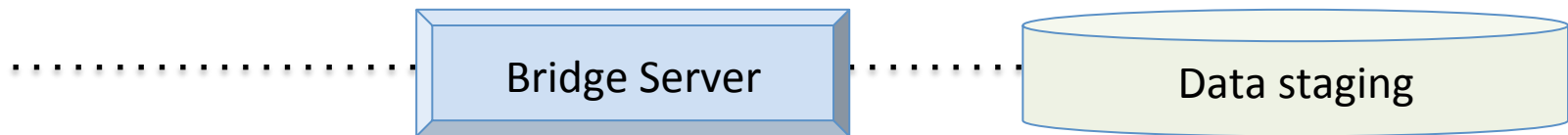
DuraCloud + Chronopolis

Goals:

- Integrate the DuraCloud suite of cloud services with Chronopolis as a DPN node
- Create a node that will provide a simple web-based upload to DPN
- Provide an option that combines cloud-based access with DPN preservation

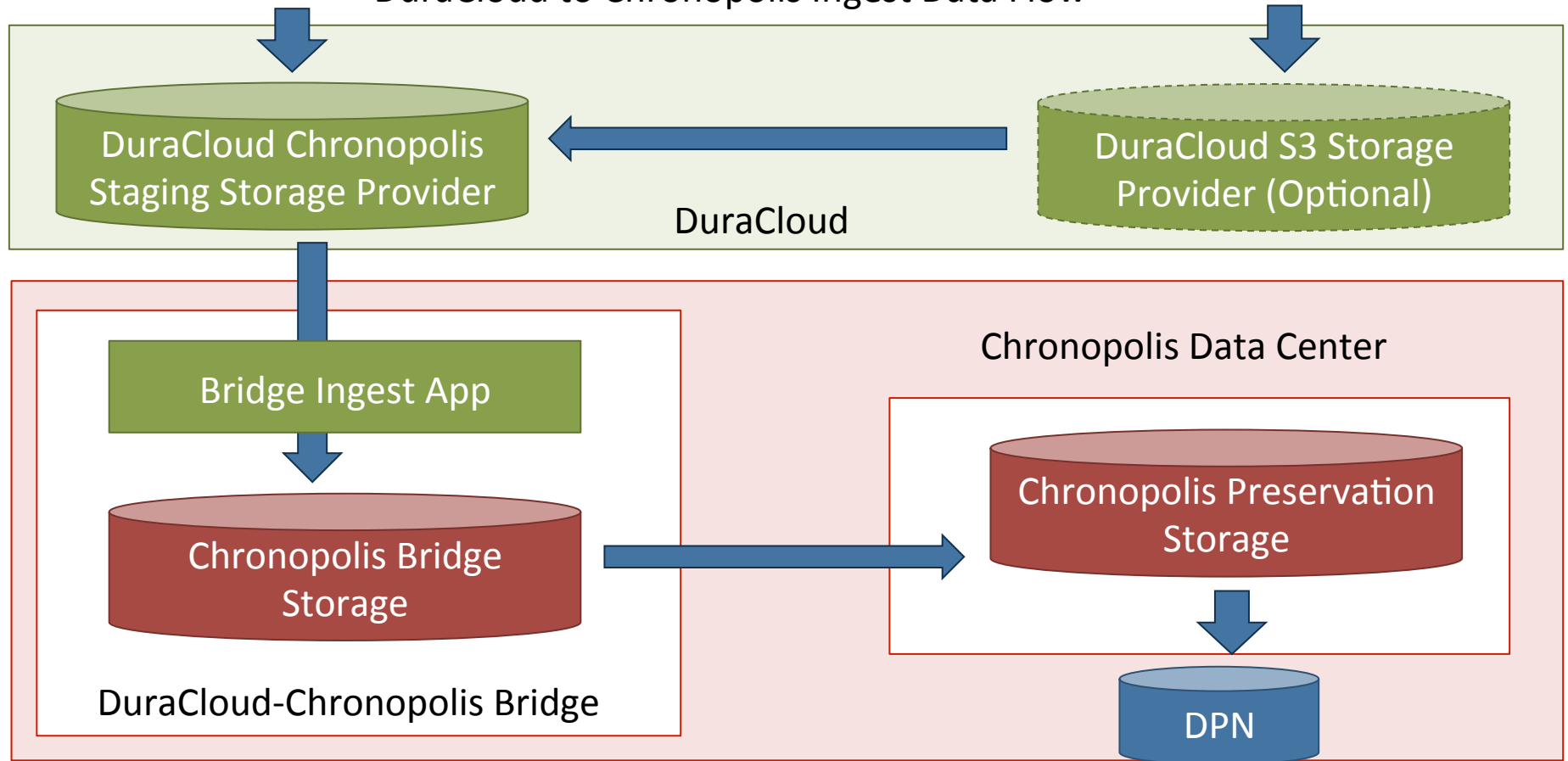


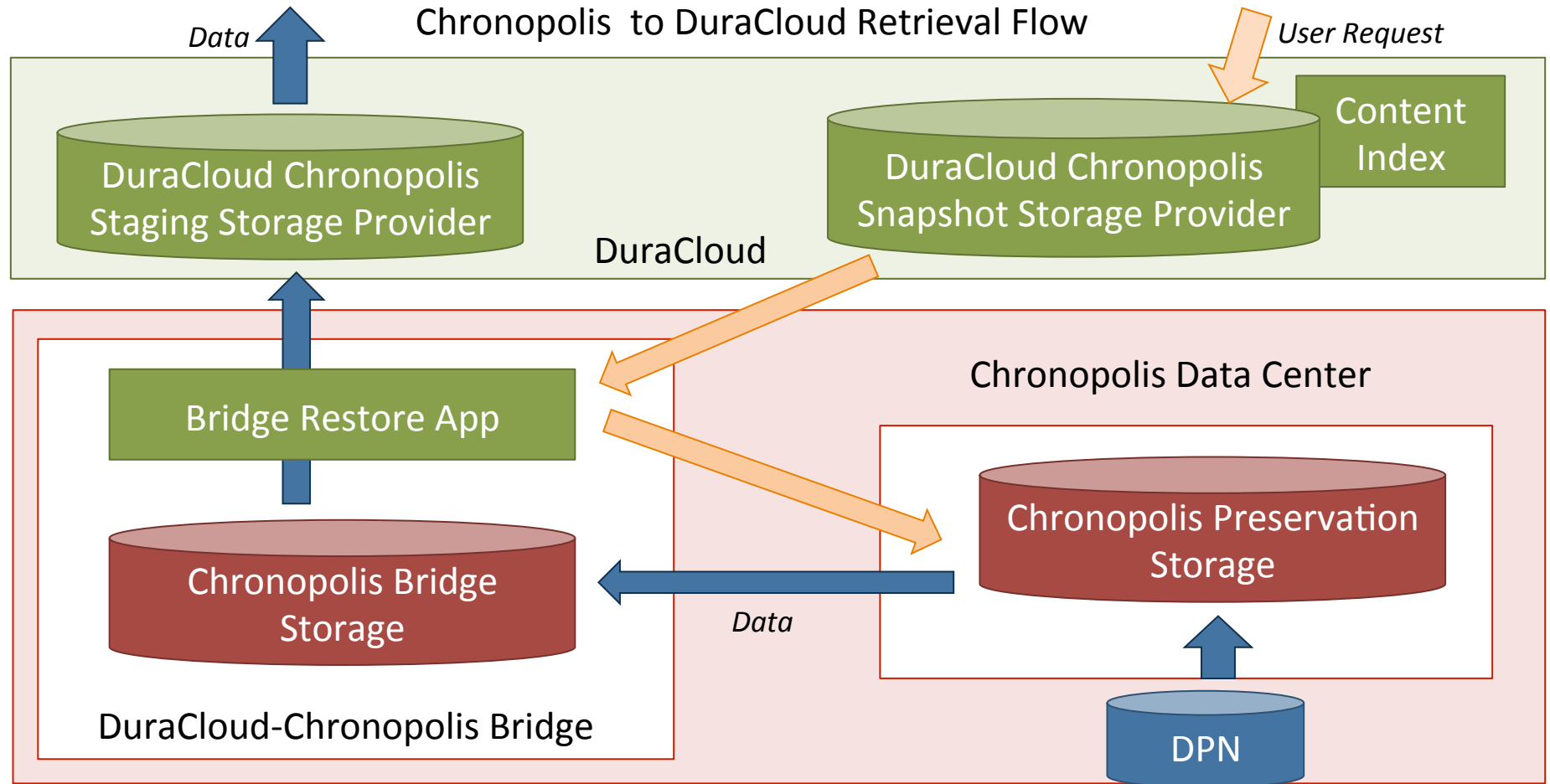
DuraCloud
Infrastructure



Chronopolis
Infrastructure

DuraCloud to Chronopolis Ingest Data Flow





Texas Preservation Node DSpace/DPN Integration



University of Texas Libraries



Introduction

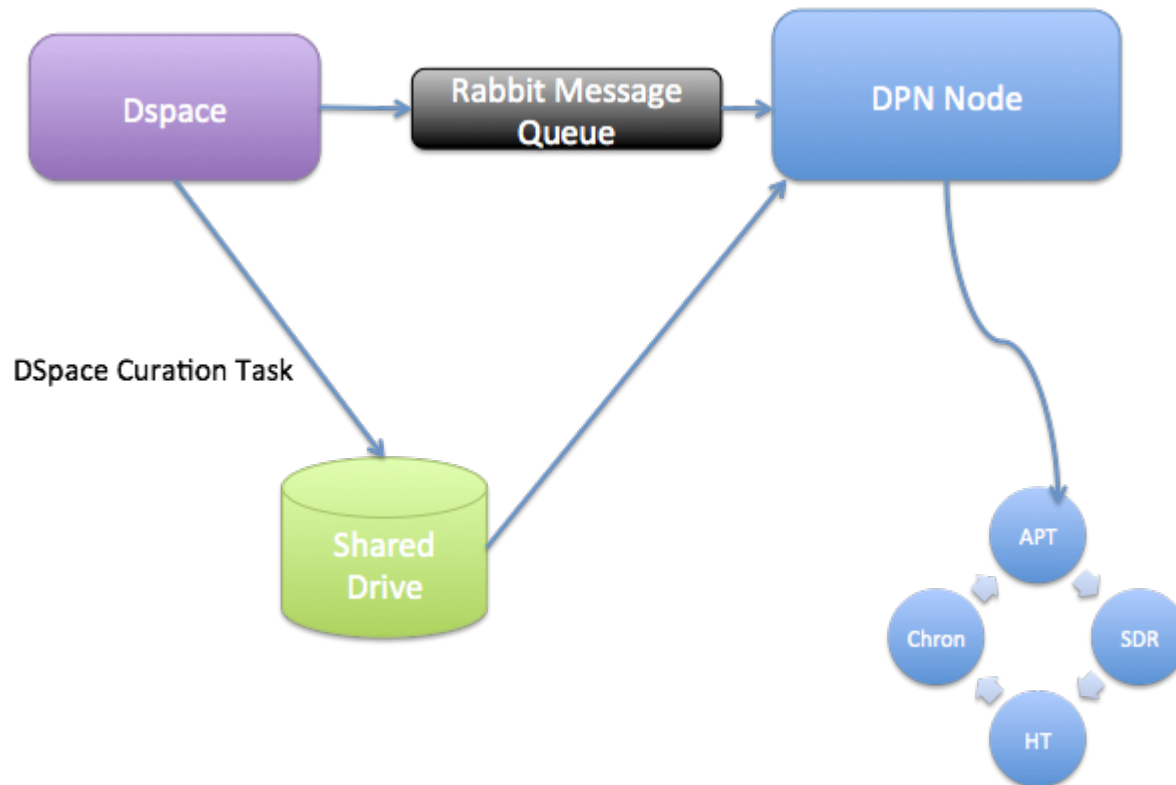
- The Texas Preservation Node is a collaboration between the University of Texas Library (UTL), the Texas Digital Library (TDL) and the Texas Advanced Computing Center (TACC).
- The UTL and the TDL use DSpace as their institutional repository.
- We wanted to explore mechanisms for moving assets from the institutional repository into the DPN network for preservation.
- We wanted to use the web UI built into DSpace.

The Process - Export from DSpace

- We used the DSpace curation system to enable an end user to enter the handle of a collection that was to be sent to DPN.
- A DSpace curation task was written that took the handle, exported the assets and notified the DPN node.
- The assets were exported to an NFS volume that was shared between the DSpace server and the DPN node

The Process - Stage for DPN

- After the assets were exported, a message was sent via a RabbitMQ message queue to the DPN node.
- There was a process running on the DPN node - waiting for messages.
 - The process creates a DPN bag from the assets (from the shared NFS volume)
 - Puts a copy of the DPN bag at TACC via iRODS
 - Puts a copy of the DPN bag in the DPN staging area
 - Initiates the transmission of the bag into DPN



Conclusion

- The DSpace curation system provided a good way to allow an end user to indicate a collection for transmission to DPN.
- Using NFS and RabbitMQ made it pretty straightforward to move content to the DPN ingestion point and to kick off the process of ingesting content into DPN.
- A technique similar to this could be used for moving assets from any repository into DPN.

Stanford Digital Repository (SDR)-DPN Integration



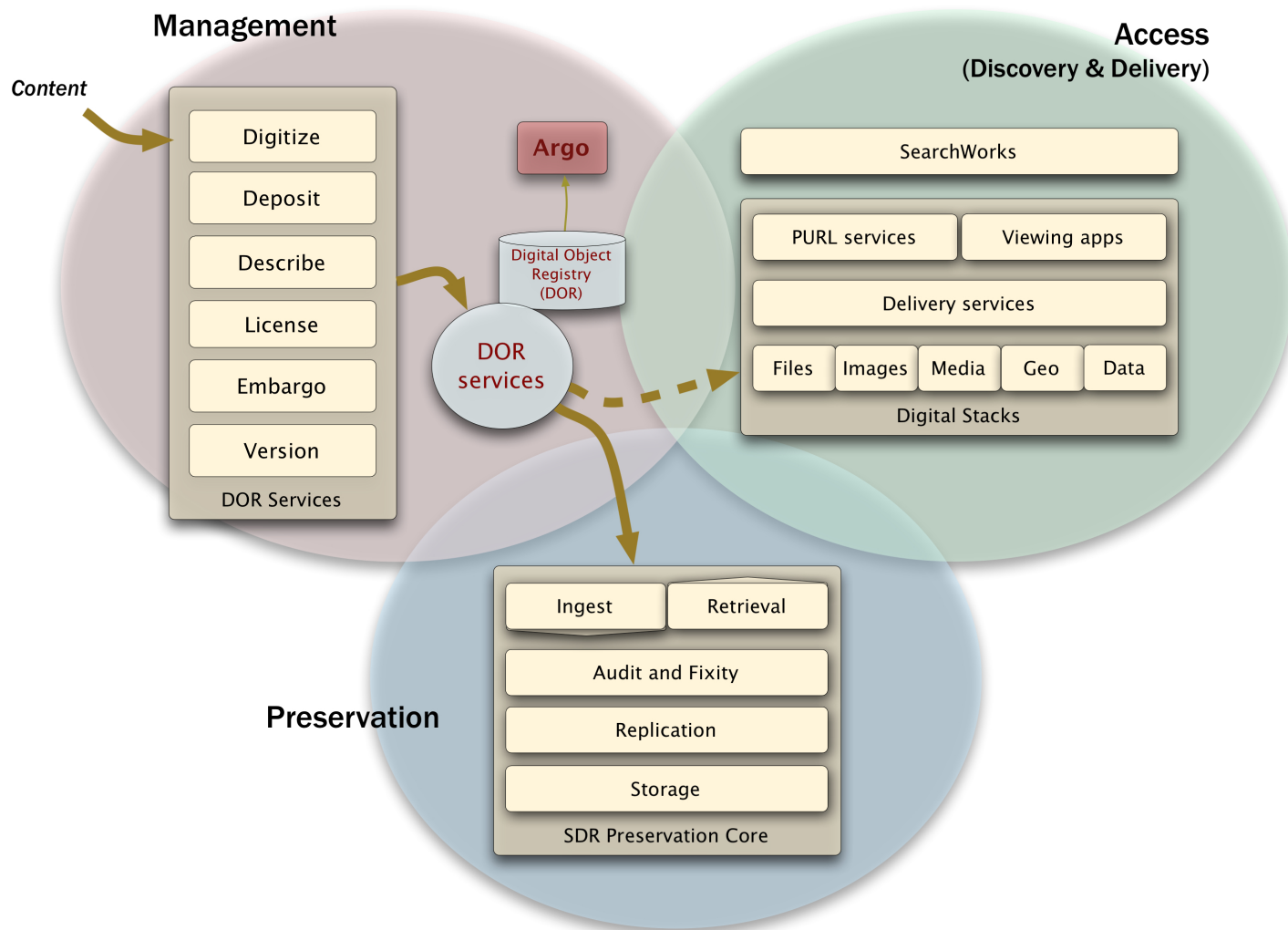
STANFORD UNIVERSITY LIBRARIES

The Stanford Digital Repository (SDR) is...

- a Fedora-based preservation repository,
- with Hydra-based management and access services,
- a modular architecture,
- and serves as a DPN Node.

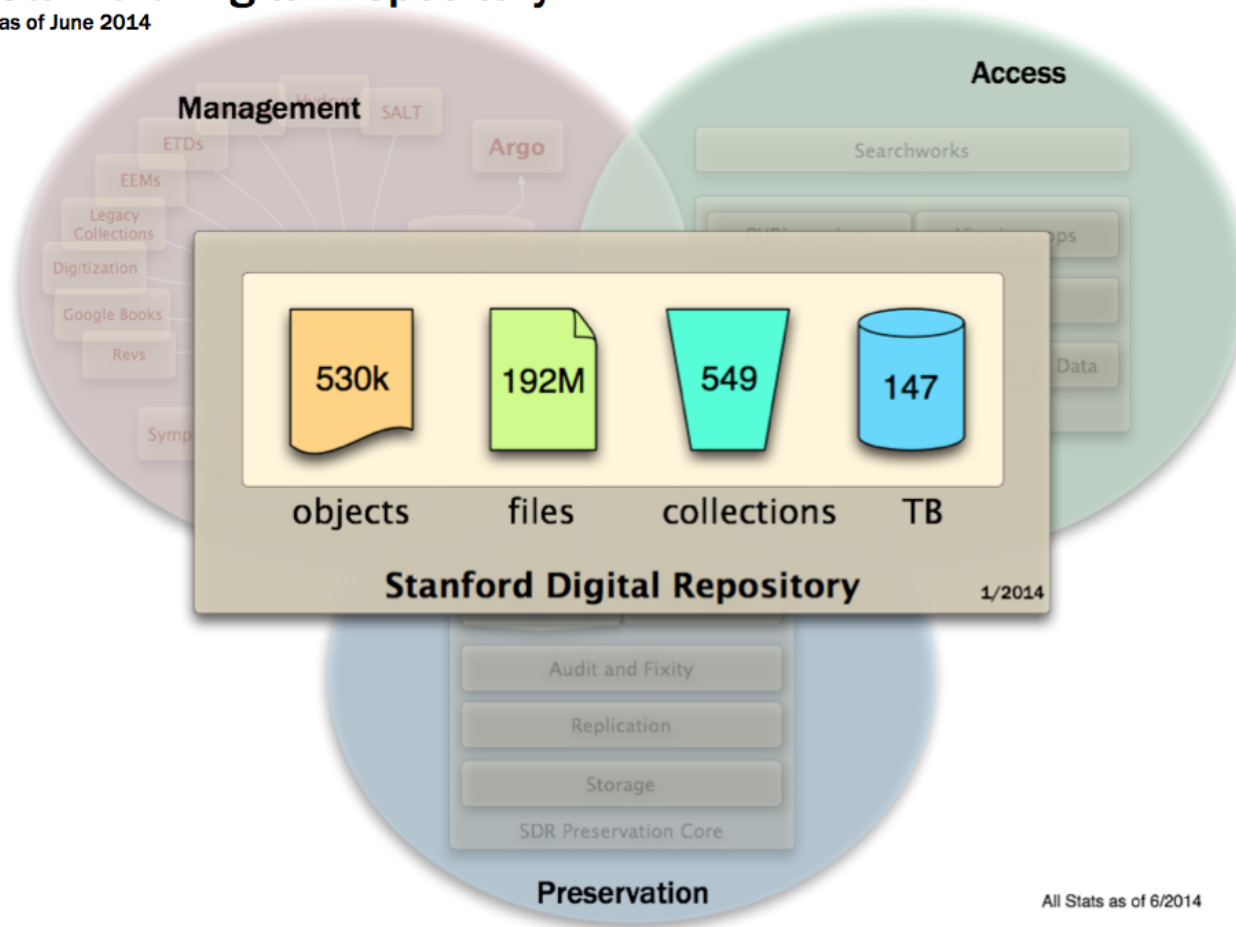


<http://library.stanford.edu/sdr>

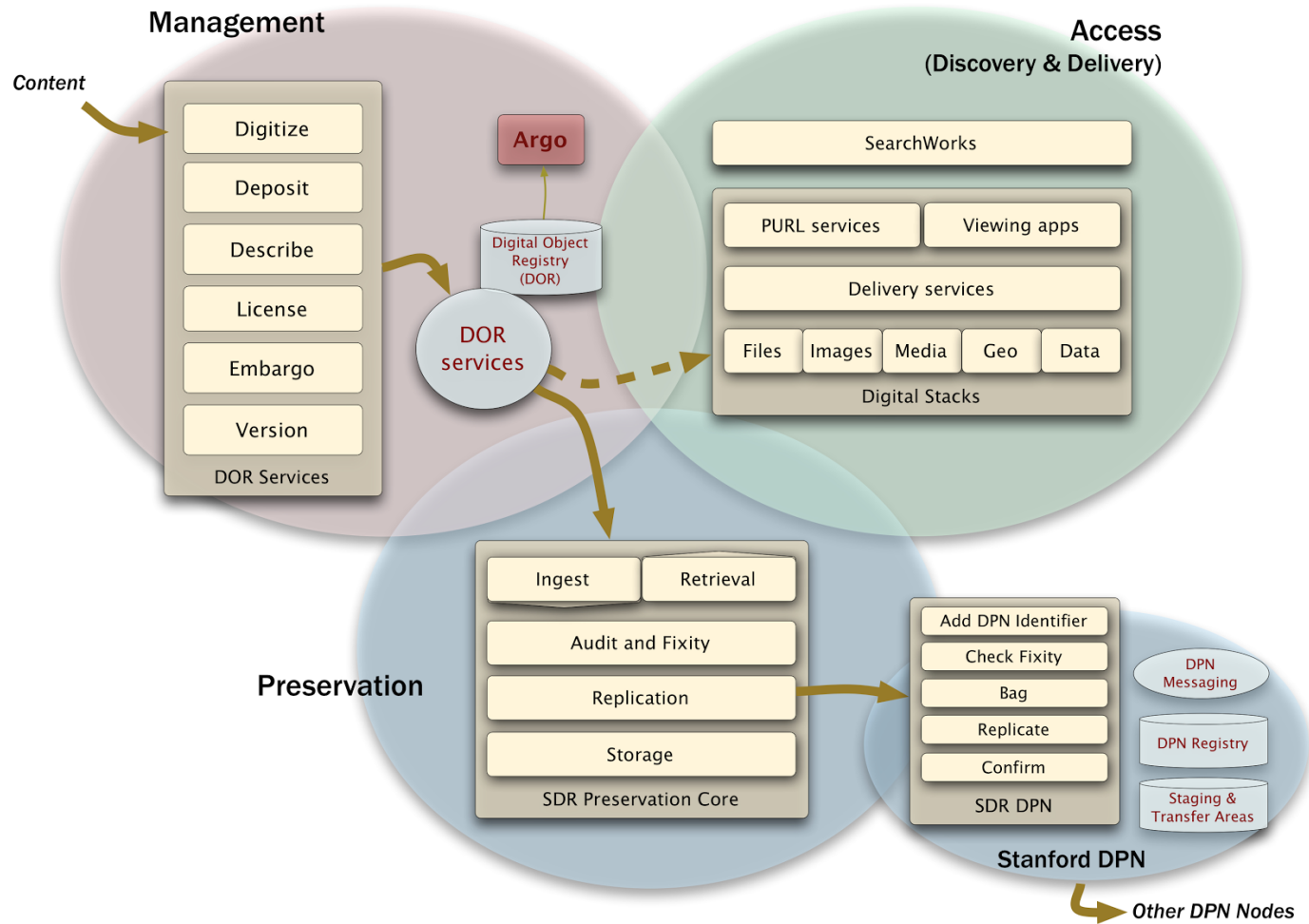


The Stanford Digital Repository

Contents as of June 2014



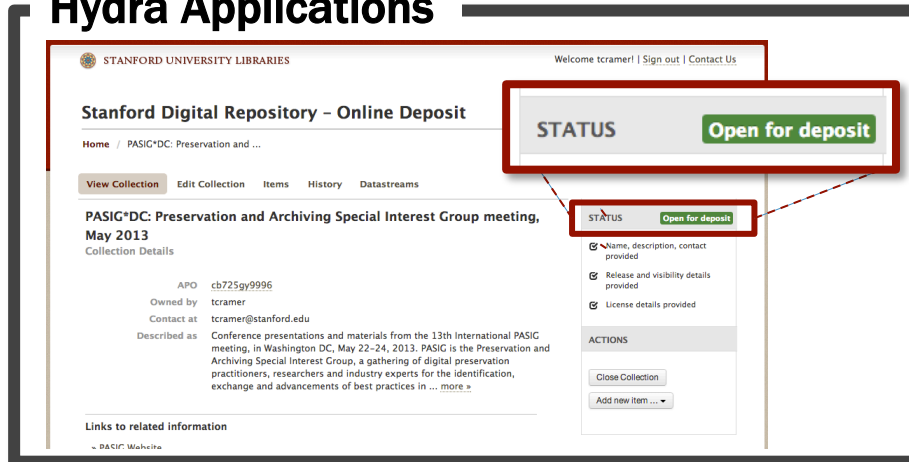
All Stats as of 6/2014



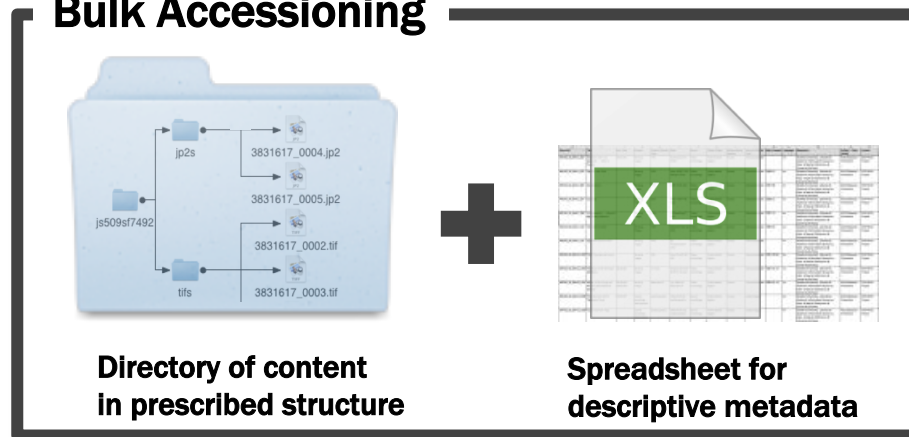
All SDR Content is covered by an Agreement

- All deposits are governed by exactly one set of agreements governing content administration
 - Encapsulates contracts, MOUs and terms of deposit
 - Defines owners, roles, timeframe, preservation and access services
 - *Will include DPN agreements for replication and succession rights*
- Agreement terms are recorded in machine-actionable form in an Administrative Policy Object (APO)
 - Every object isGovernedBy one APO
 - Records default <rights, access, attribution, storage> settings for collections
 - “Replicate to DPN” flag will be set in an APO

Hydra Applications



Bulk Accessioning

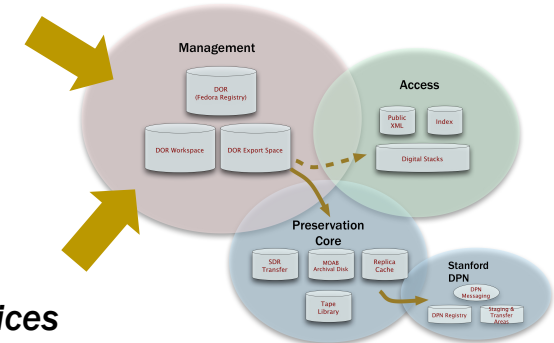


Pathways into SDR

Hydra apps create well-formed objects in Fedora

-AND/OR-

Accessioning services transform content from directory, build metadata data streams from XLS and directory, and assemble full objects in Fedora via bulk loading.



Augmenting Hydra for DPN

Terms of Use and Licenses

The following Terms of Use will be displayed with each item published in this collection: "User agrees that, where applicable, content will not be used to identify or to otherwise infringe the privacy or confidentiality rights of individuals. Content distributed via the Stanford Digital Repository may be subject to additional license and use restrictions applied by the depositor."

After its release date, items in this collection should be covered by:

- ☒ no license
- ☐ varies -- contributor may select a license for each item, with a default of
- ☐ required license -- applies to all items in the collection

Collection participants

Provide a SUNet ID (NOT email address) for each person who can manage, deposit or view items in this collection. Separate multiple IDs with a comma. You only need to add a user once for the highest level role, since they will have the ability to perform the actions for any roles below.

Owner
This is the owner of the collection, and cannot be changed

Managers
These users can edit collection details, and add and review items in the collection

Depositors
These users can add items to the collection, but cannot act as reviewers

Viewers
These users can view items in the collection only

Review Workflow

Select the optional review workflow if you want each submitted item to be reviewed before final deposit into the repository.

- ☒ No

STATUS

Closed for deposit

- ☐ Name, description, contact provided
- ☒ Release and visibility details provided
- ☒ License details provided
- ☐ Collection can be opened for deposit

ACTIONS

- ☒ Replicate collection to DPN
I agree to the [DPN terms of service](#)

S

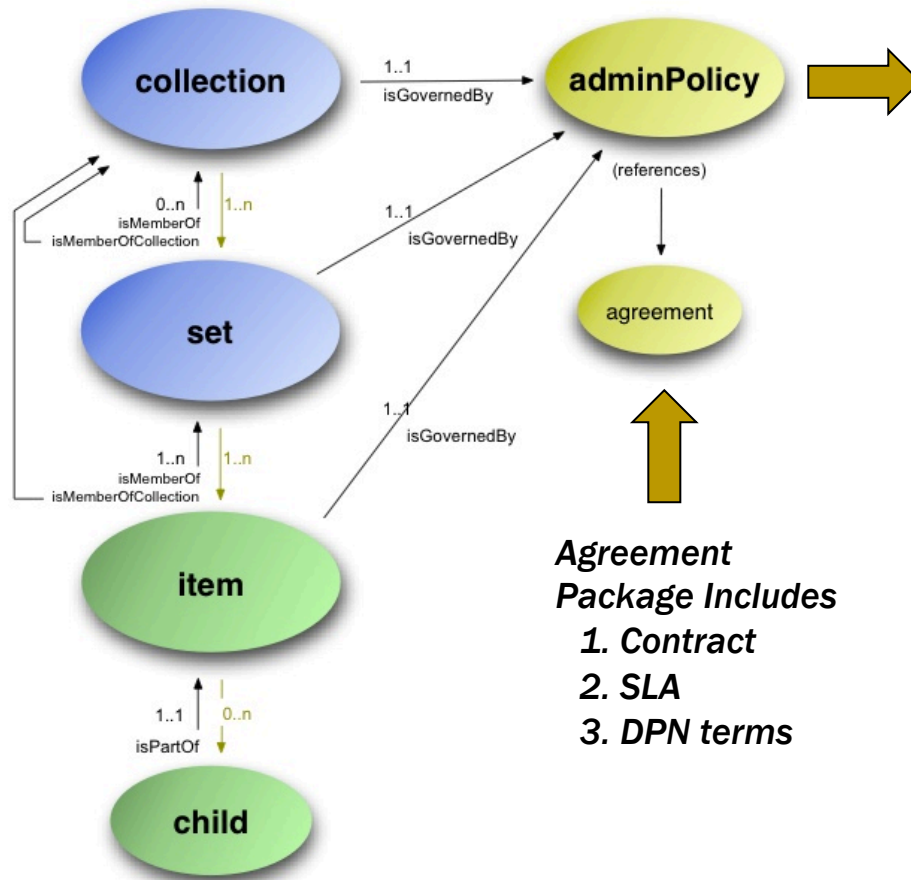
ACTIONS

- ☒ Replicate collection to DPN

I agree to the [DPN terms of service](#)

Save

DPN Service Will Be Declared in the APO



Administrative Policy Object (APO) contains an administrativeMetadata data stream

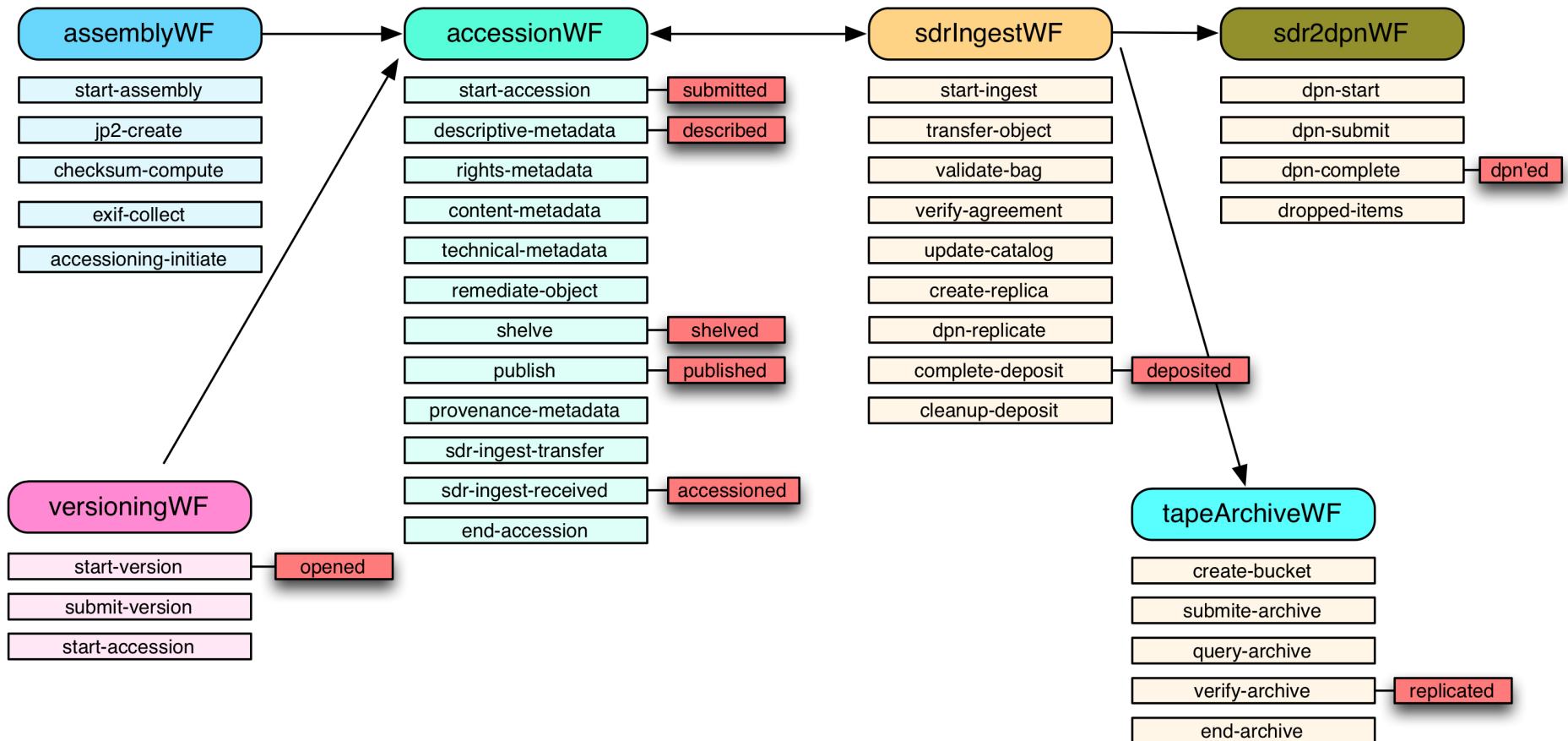
```

<administrativeMetadata objectId="druid:...">
  <descMetadata>
    <format>MODS</format>
    <source>symphony</source> </descMetadata>
  ....
  <accessioning>
    <shelve value="yes" mimetype="image/jp2" />
    <publish value="yes" mimetype="image/jp2" />
    <preserve value="yes" mimetype="*" />
  </accessioning>
  <preservation>
    <disk copy value="yes" mimetype="*" />
    <tape copy value="yes" mimetype="*" />
    <replicate to dpn="yes" mimetype="*" />
  </preservation>
  <dissemination>
    <harvester name="searchworks">item</harvester>
    <harvester name="searchworks">collection</harvester>
    <releaseDelayLimit>20 years</releaseDelayLimit>
  </dissemination>
  ....
</administrativeMetadata>
  
```

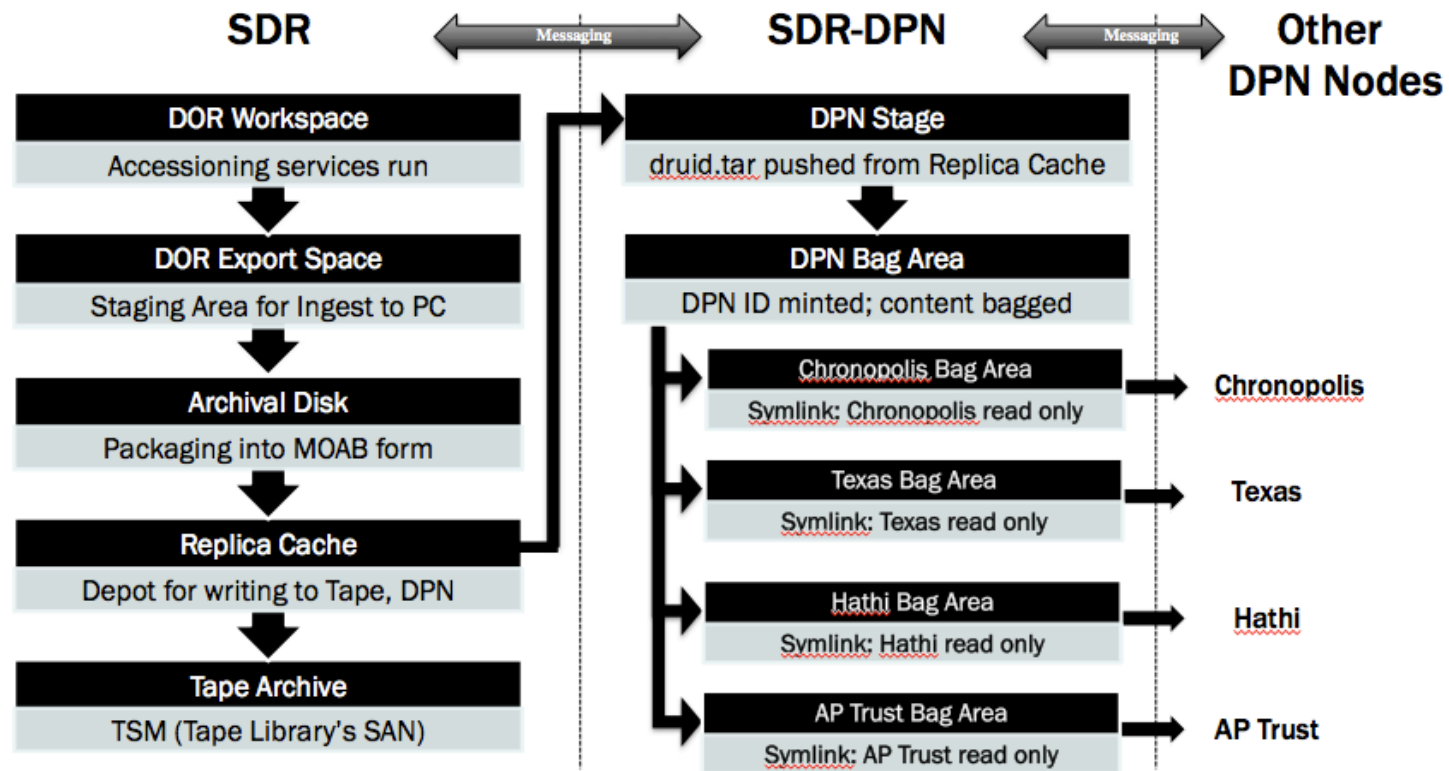
Agreement Package Includes

1. Contract
2. SLA
3. DPN terms

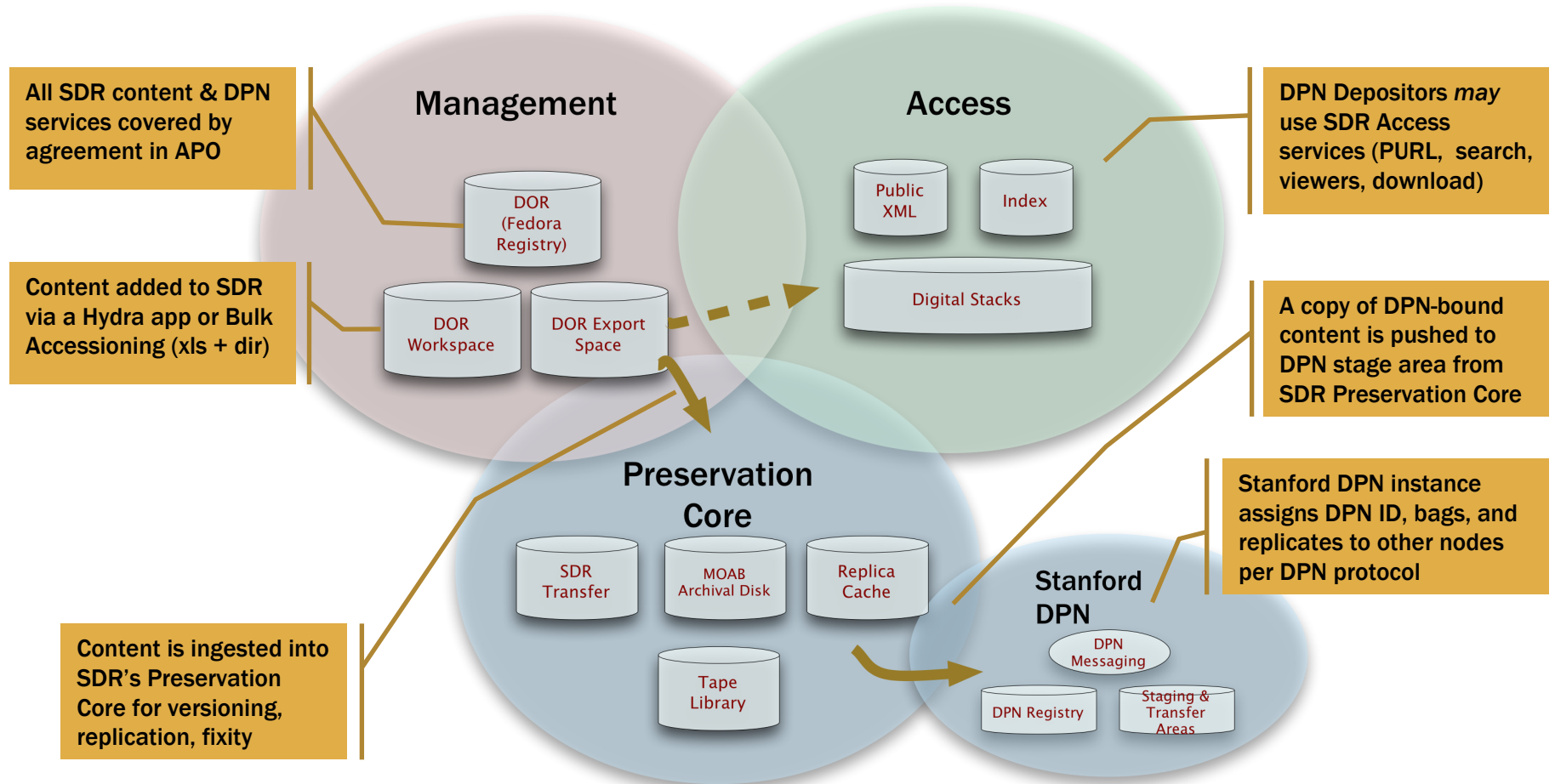
SOA: Content Moves Across SDR via Workflows



Content Movement From SDR to DPN



Synopsis: SDR 2 DPN



OH BY THE WAY



- DPN pilots starting in 2014
- DPN is **HIRING** – job postings coming up!

Thank you!



dminor@ucsd.edu

dan.galewsky@austin.utexas.edu

tcramer@stanford.edu

inquiry@dpn.org
<http://www.dpn.org>